An Extended Fano's Inequality for the Finite Blocklength Coding

Yunquan Dong, Pingyi Fan
{dongyq08@mails,fpy@mail}.tsinghua.edu.cn
Department of Electronic Engineering, Tsinghua University, Beijing, P.R. China.

***Abstract*—Fano's inequality reveals the relation between the conditional entropy and the probability of error . It has been the key tool in proving the converse of coding theorems in the past sixty years. In this paper, an extended Fano's inequality is proposed, which is tighter and more applicable for codings in the finite blocklength regime. Lower bounds on the mutual information and an upper bound on the codebook size are also given, which are shown to be tighter than the original Fano's inequality. Especially, the extended Fano's inequality is tight for some symmetric channels such as the q-ary symmetric channels (QSC).**

***Index Terms*—Fano's inequality, finite blocklength regime, channel coding, Shannon theory.**

I. Introduction

As known to all, Shannon's information theory deals mainly with the representation and transmission of information. In the development of both source and channel coding theorems, especially for their converses, Fano's inequality serves as the key tool [2].

Theorem 1: Fano's Inequality

X and Y are two random variables following $(X, Y) \backsim p(x, y)$ and $X, Y \in \mathcal{X}$. Define $P_e = \Pr\{X \neq Y\}$, then

$$H(X|Y) \leq H(P_e) + P_e \log(M-1) \tag{1}$$

where $M = |\mathcal{X}|$ is the cardinality of X and Y, $H(P_e)$ is the binary entropy function $H(x) = -[x\log x + (1-x)\log(1-x)]$ for $0 \leq x \leq 1$. Thus Fano' inequality can be further relaxed by $H(P_e) \leq 1$.

Usually, the left hand side of (1) is referred to as the equivocation, which is quantified by the conditional entropy. Particularly, it represents the uncertainty whether the restored/decoded message Y is the same as the original one, i.e., X. On the other hand, the right hand side implies the reliability of the source/channel coding in terms of a function of error probability P_e. It was shown in [3] that vanishing equivocation implies vanishing error probability. However, vanishing error probability does not necessarily guarantee a vanishing equivocation, especially for some X, Y of countably infinite alphabet.

In proving the converse of coding theorems, one wants to find the upper bound on the size of the codebook given arbitrary code length and error probability. The following theorem is an immediate inference of Fano's inequality, simple but useful.

Theorem 2: [4] Suppose X and Y are two random variables that take values on the same finite set with cardinality M and at least one of them is equiprobable. Then the mutual information between them satisfies

$$I(X;Y) \geq (1-P_e)\log M - H(P_e), \tag{2}$$

where $P_e = \Pr\{X \neq Y\}$ and $H(x) = -[x\log x + (1-x)\log(1-x)]$.

As an inference of this result, the following theorem gives an upper bound on the size of a code as a function of the average error probability.

Theorem 3: [8] Every (M, ϵ)-code (average probability of error) for a random transformation $P_{Y|X}$ satisfies

$$\log M \leq \frac{1}{(1-\epsilon)}\sup_X I(X;Y) + \frac{1}{(1-\epsilon)}H(\epsilon), \tag{3}$$

where $\epsilon = \Pr\{X \neq Y\}$, $H(x) = -[x\log x + (1-x)\log(1-x)]$.

Although simple, these two theorems are insightful and easy to compute both in theory and numerically.

The classical coding theorems are mainly based on the asymptotic equipartition property (AEP) and typical/joint-typical decoder [1]. However, applications of AEP requires infinite long codewords, where the error probability goes either to 0 or 1 as the code length goes to infinity. Although these coding theorems provide fundamental limits for modern communications, research on the finite blocklength coding schemes are more important in engineering applications. Given the block length, upper bounds on the achievable error probability and the achievable code size were obtained in [8]. Most importantly, a tight approximation for the achievable maximal rate given the error probability and code length was presented.

In this paper, we consider the entropy of one random variable vector conditioned on another, and the corresponding probability of error in guessing one from the other, by proposing an extended Fano's inequality. The extended Fano's equality has better performance by taking advantage of a more careful consideration on the error patterns. It suits codings in the finite blocklength regime better and is useful in bounding the mutual information between random vectors, and the codebook size given the block length and average symbol error probability constraint.

In the following part of this paper, we present the extended Fano's inequality in Section II first. The lower bounds on the mutual information between two random variable vectors and a upper bound on the codebook size given the block length and error probability are given in Section III. An application of the

the obtained result to the q-ary symmetric channels (QSC) are presented in Section IV, which shows that the extended Fano's inequality is tight for such channels. Finally, we concluded the paper in Section V. Throughout the paper, vectors indicated by bold.

II. Fano's Inequality Extension

Although Fano's inequality has been used widely in the past few years, it can be improved by treating the error events more carefully. In this section, a refinement of Fano's inequality is presented, which is tighter and more applicable for finite blocklength coding design.

Theorem 4: Fano's Inequality Extension

Suppose that $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ and $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ are two n-dimension random vectors where X_k and Y_k $(k = 1, 2, \cdots, n)$ take values on the same finite set $\mathcal{X}$ with cardinality $|\mathcal{X}| = q$. Then the conditional entropy satisfies

$$H(\boldsymbol{X}|\boldsymbol{Y}) \le H(\boldsymbol{p}) + \sum_{k=1}^{n} p_k \log \left(C_n^k (q-1)^k \right), \tag{4}$$

where $H(\boldsymbol{p}) = -\sum_{k=0}^{n} p_k \log p_k$ is the discrete entropy function. $\boldsymbol{p} = \{p_0, p_1, \cdots, p_n\}$ is the error distribution, where the error probabilities are $p_k = \Pr\left(H_d(\boldsymbol{X}, \boldsymbol{Y}) = k\right)$ for $k = 0, 1, \cdots, n$. $H_d(\boldsymbol{X}, \boldsymbol{Y})$ is the generalized Hamming distance, defined as the number of symbols in $\boldsymbol{X}$ that are different from the corresponding symbol in $\boldsymbol{Y}$.

Proof: Define the error random variable as $E = k$ if $H_d(\boldsymbol{X}, \boldsymbol{Y}) = k$ for $k = 0, 1, \cdots, n$. According to the chain rule of the joint entropy, $H(E, \boldsymbol{X}|\boldsymbol{Y})$ can be expressed in the following ways,

$$H(E, \boldsymbol{X}|\boldsymbol{Y}) = H(\boldsymbol{X}|\boldsymbol{Y}) + H(E|\boldsymbol{X}, \boldsymbol{Y}) \tag{5a}$$

$$= H(E|\boldsymbol{Y}) + H(\boldsymbol{X}|E, \boldsymbol{Y}) \tag{5b}$$

Particularly, in (5.a), it is clear that $H(E|\boldsymbol{X}, \boldsymbol{Y}) = 0$. Then we have

$$\begin{aligned} H(\boldsymbol{X}|\boldsymbol{Y}) =& H(E|\boldsymbol{Y}) + H(\boldsymbol{X}|E, \boldsymbol{Y}) \\ \overset{(a)}{\le}& H(E) + H(\boldsymbol{X}|E, \boldsymbol{Y}), \end{aligned} \tag{6}$$

where (a) follows the fact that entropy increases if its condition is removed, i.e., $H(E) \ge H(E|\boldsymbol{Y})$. Particularly, we have $H(E) = H(\boldsymbol{p}) = -\sum_{k=0}^{n} p_k \log p_k$.

According to its definition, we have

$$H(\boldsymbol{X}|E, \boldsymbol{Y}) = \sum_{k=0}^{n} p_k H(\boldsymbol{X}|E = k, \boldsymbol{Y}). \tag{7}$$

When considering $H(\boldsymbol{X}|E = k, \boldsymbol{Y})$, we know that there are k disaccord symbol pairs between $\boldsymbol{X}$ and $\boldsymbol{Y}$. For each fixed $\boldsymbol{Y} = \boldsymbol{y}$, every symbol in $\boldsymbol{X}$ which belongs to a disaccord pair has $q-1$ possible choices except the one in $\boldsymbol{y}$. Thus $\boldsymbol{X}|(E = k, \boldsymbol{y})$ has $(q-1)^k$ choices. Besides, there C_n^k selections for the positions of error symbols for each given k. Therefore, the total number of possible codeword $\boldsymbol{X}$ is $C_n^k (q-1)^k$, which means

$$H(\boldsymbol{X}|E = k, \boldsymbol{Y}) \le \log \left(C_n^k (q-1)^k \right). \tag{8}$$

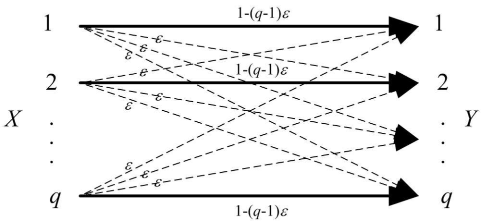

Fig. 1. The q-ary Symmetric Channel

Particularly, note that $H(\boldsymbol{X}|E = 0, \boldsymbol{Y}) = 0$ since there is no uncertainty in determining $\boldsymbol{X}$ from $\boldsymbol{Y}$, if they are the same.

Then, (7) can be written as

$$\begin{aligned} H(\boldsymbol{X}|E, \boldsymbol{Y}) =& \sum_{k=1}^{n} P(E = k) H(\boldsymbol{X}|E = k, \boldsymbol{Y}) \\ \le& \sum_{k=1}^{n} p_k \log \left(C_n^k (q-1)^k \right). \end{aligned} \tag{9}$$

By combining (6) and (9), the proof of the theorem is completed.

■

Remark 1: In fact, the error distribution $\boldsymbol{p}$ is easy to calculate, especially for some special channels. For example, the discrete q-ary symmetric channels is shown in Fig 1. In this situation, $p_k = C_n^k ((q-1)\varepsilon)^k (1 - (q-1)\varepsilon)^{n-k}$.

Remark 2: It is clear that Theorem 4 is a generalization of Fano's inequality. Specifically, when the block length is 1, i.e., $n = 1$, we have $p_1 = p_e$, $p_0 = 1 - p_e$, and $C_1^1 = 1$. In this case, Theorem 4 reduces to

$$H(X|Y) \le H(p_e) + p_e \log(q-1) \tag{10}$$

which is exactly the same as Fano's inequality.

As a variant of Theorem 4, the following theorem presents the conditional entropy in terms of relative entropy.

Theorem 5: Suppose that $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ and $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ are two n-dimension random vectors where X_k and Y_k $(k = 1, 2, \cdots, n)$ take values on the same finite set $\mathcal{X}$ with cardinality $|\mathcal{X}| = q$. Then the conditional entropy satisfies

$$H(\boldsymbol{X}|\boldsymbol{Y}) \le n \log q - D(\boldsymbol{p}\|\boldsymbol{q}), \tag{11}$$

where $D(\boldsymbol{p}\|\boldsymbol{q}) = \sum_{k=0}^{n} p_k \log \frac{p_k}{q_k}$ is the discrete relative entropy function. The error probabilities are $p_k = \Pr\left(H_d(\boldsymbol{X}, \boldsymbol{Y}) = k\right)$ for $k = 0, 1, \cdots, n$. Donate $\boldsymbol{p} = \{p_0, p_1, \cdots, p_n\}$, and $\boldsymbol{q} = \{q_0, q_1, \cdots, q_n\}$ is a probability distribution with $q_k = \frac{C_n^k (q-1)^k}{q^n}$

Proof: Firstly, we know from the binomial theorem that $\{\frac{C_n^k a^k b^{n-k}}{(a+b)^n}\}_{0 \le n}$ is a probability distribution. Let $a = q-1$ and $b = 1$, we know that $\boldsymbol{q}$ is a probability distribution where $q_k = \frac{C_n^k (q-1)^k}{q^n}$.

According to Theorem 4,

$$H(\boldsymbol{X}|\boldsymbol{Y}) \overset{(a)}{\le} -\sum_{k=0}^{n} p_k \log \frac{p_k}{C_n^k (q-1)^k} = n \log q - D(\boldsymbol{p} \parallel \boldsymbol{q})$$

where (a) holds because $\log\left(C_n^k(q-1)^k\right)=0$ for $k=0$. ■

Remark 3: By the definition of $\boldsymbol{p}$, it is clear that it reflects the error performance of the channel and is totaly determined by the channel itself. On the contrary, $\boldsymbol{q}$ is a distribution where each error patten is assumed to appear equiprobablely. In this situation, the probability that there are k error symbols in the codeword is $q_k = C_n^k(\frac{q-1}{q})^k(\frac{1}{q})^{n-k}$. Thus $D(\boldsymbol{p}\|\boldsymbol{q})$ is the distance between the actual error pattern distribution and the uniform error pattern distribution. Particularly, if the channel is an error free one, i.e., $p_0=1$ and $p_k=0$ for $1\le k\le n$, we have $D(\boldsymbol{p}\|\boldsymbol{q}) = \sum_{k=0}^n p_k\log\frac{p_k}{q_k} = 1\cdot\log\frac{1}{q_0} = \log q^n$. According to Theorem 5, we get $H(\boldsymbol{X}|\boldsymbol{Y})=0$, which is reasonable with the assumption on the channel. In this sense, Theorem 5 is tight.

The Fano's inequality has been playing an important role in the history of information theory because it built a close connection between conditional entropy and error probability. For extended Fano's inequality given in Theorem 4, it is especially applicable in finite blocklength coding. It also presents the relationship between conditional entropy and error probabilities, which are defined as follows.

Definition 1: Block error probability P_b is the average error probability of a block (codeword), i.e., $P_b=\Pr\{\boldsymbol{X}\neq\boldsymbol{Y}\}$. Then we have $P_b=\sum_{k=1}^n p_k$. Thus, we have $P_b\ge p_k$ for any $0<k\le n$.

Definition 2: Symbol error probability P_s is the average error probability of a symbol, i.e., $P_s=\Pr\{X_k\neq Y_k\}$, which can be expressed by $P_s=\frac{1}{n}\sum_{k=1}^n kp_k$.

Remark 4: In many communication systems, especially those using error correction channel codings, a block error doesn't imply a system failure. On the contrary, the error can be corrected or part of the block can still be used with some performance degradation. In this case, the symbol error is more useful than the block error.

Particularly, a corollary following our result as shown below will answer this problem.

Corollary 1: Suppose that $\boldsymbol{X}=\{X_1,X_2,\cdots,X_n\}$ and $\boldsymbol{Y}=\{Y_1,Y_2,\cdots,Y_n\}$ are two n-dimension random vectors where X_k and Y_k take values on the same finite set $\mathcal{X}$ with cardinality $|\mathcal{X}|=q$. Then the conditional entropy satisfies

$$H(\boldsymbol{X}|\boldsymbol{Y})\le n-D(\boldsymbol{p}\|\boldsymbol{w})+nP_s\log(q-1), \tag{12}$$

where $\boldsymbol{w}=\{w_0,w_1,\cdots,w_n\}$ is a probability distribution with $w_k=\frac{C_n^k}{2^n}$.

Proof: According to Theorem 4, one has

$$\begin{aligned}H(\boldsymbol{X}|\boldsymbol{Y}) &\le H(\boldsymbol{p})+\sum_{k=0}^n p_k\log\left(\frac{C_n^k}{2^n}2^n\right)+\sum_{k=1}^n p_k\log(q-1)^k\\ &=-\sum_0^n p_k\log p_k+\sum_{k=0}^n p_k\log w_k+\sum_{k=0}^n p_k\log 2^n\\ &=n-D(\boldsymbol{p}\|\boldsymbol{w})+nP_s\log(q-1).\end{aligned}$$

■

Remark 5: Since the distribution $\boldsymbol{w}$ can be expressed as $w_k=C_n^k(\frac{1}{2})^k(\frac{1}{2})^{n-k}$, which is a binomial distribution with the symbol error probability of 0.5. $D(\boldsymbol{p}\|\boldsymbol{w})$ is a measure of the distance between the error probability distribution and the binomial distribution with parameter 0.5.

Remark 6: If one takes $n=1$, Corollary 1 will reduces to $H(X|Y)\le 1+P_s\log(q-1)$, which is a frequently used form of Fano's inequality.

Remark 7: It is seen that the extended Fano's inequality builds a natural connection between conditional entropy and symbol error and is especially applicable for finite length codings.

III. Converse Results

A. Lower Bounds on the Mutual Information

Based on the proposed generalized Fano's inequality, the following lower bounds on the mutual information between $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be obtained.

Theorem 6: Suppose that $\boldsymbol{X}=\{X_1,X_2,\cdots,X_n\}$ and $\boldsymbol{Y}=\{Y_1,Y_2,\cdots,Y_n\}$ are two n-dimension random vectors that satisfy the following.

1) X_k and Y_k $(k=1,2,\cdots,n)$ take values on the same finite set $\mathcal{X}$ with cardinality $|\mathcal{X}|=q$.
2) Either $\boldsymbol{X}$ or $\boldsymbol{Y}$ is equiprobable.
3) The error probabilities are $p_k=\Pr\left(H_d(\boldsymbol{X},\boldsymbol{Y})=k\right)$ for $k=0,1,\cdots,n$. Donate the error distribution as $\boldsymbol{p}=\{p_0,p_1,\cdots,p_n\}$.

Then the mutual information between $\boldsymbol{X}$ and $\boldsymbol{Y}$ satisfies

$$I(\boldsymbol{X};\boldsymbol{Y})\ge n\log q-H(\boldsymbol{p})-\sum_{k=1}^n p_k\log\left(C_n^k(q-1)^k\right), \tag{13}$$

where $H(\boldsymbol{p})=-\sum_{k=0}^n p_k\log p_k$ is the entropy function.

Proof: If $\boldsymbol{X}$ is equiprobable, $H(\boldsymbol{X})=\log q^n=n\log q$. On the other hand, the mutual information is given by $I(\boldsymbol{X};\boldsymbol{Y})=H(\boldsymbol{X})-H(\boldsymbol{X}|\boldsymbol{Y})$. Together with Theorem 4,

$$I(\boldsymbol{X};\boldsymbol{Y})\ge n\log q-H(\boldsymbol{p})-\sum_{k=1}^n p_k\log\left(C_n^k(q-1)^k\right). \tag{14}$$

Note that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are totally symmetric in (14). Therefore, if $\boldsymbol{Y}$ is assumed to be equiprobable at the beginning of the proof, one can get the same result. Thus Theorem 6 is proved. ■

By using Theorem 5, the mutual information between $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be bounded by the following Corollary.

Corollary 2: Suppose that $\boldsymbol{X}=\{X_1,X_2,\cdots,X_n\}$ and $\boldsymbol{Y}=\{Y_1,Y_2,\cdots,Y_n\}$ are two n-dimension random vectors that satisfy the following.

1) X_k and Y_k $(k=1,2,\cdots,n)$ take values on the same finite set $\mathcal{X}$ with cardinality $|\mathcal{X}|=q$.
2) Either $\boldsymbol{X}$ or $\boldsymbol{Y}$ is equiprobable.
3) The error probabilities are $p_k=\Pr\left(H_d(\boldsymbol{X},\boldsymbol{Y})=k\right)$ and $\boldsymbol{p}=\{p_0,p_1,\cdots,p_n\}$.

Then the mutual information between $\boldsymbol{X}$ and $\boldsymbol{Y}$ satisfies

$$I(\boldsymbol{X};\boldsymbol{Y})\ge D(\boldsymbol{p}\|\boldsymbol{q}) \tag{15}$$

where $D(\boldsymbol{p}\|\boldsymbol{q})=\sum_{k=0}^n p_k\log\frac{p_k}{q_k}$ is the discrete relative entropy function and $\boldsymbol{q}=\{q_0,q_1,\cdots,q_n\}$ is a probability distribution with $q_k=\frac{C_n^k(q-1)^k}{q^n}$

The distribution $\boldsymbol{q}$ means that the symbol in $\boldsymbol{Y}$ takes any value on $\mathcal{B}$ with equal probability, regardless of what is sent in $\boldsymbol{X}$. So it is a pure random distribution when X and Y are independent from each other. The most desirable coding scheme is that its error distribution $\boldsymbol{p}$ is farthermost from $\boldsymbol{q}$, which also ensures a larger coding rate.

B. Upper Bounds on the Codebook Size

Suppose $\mathcal{X}$ is a finite alphabet with cardinality $|\mathcal{X}| = q$. Let's consider the input and output alphabets $\mathcal{A}^n = \mathcal{B}^n \subseteq \mathcal{X}^n$ with $|\mathcal{A}^n| = |\mathcal{B}^n| = M$ and a channel to be a sequence of conditional probabilities [5] $\{P_{\boldsymbol{Y}|\boldsymbol{X}} : \mathcal{A}^n \mapsto \mathcal{B}^n\}$. We donate a codebook with M codewords by $(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_M) \in \mathcal{A}^n$. A decoder is a random transformation $P_{Z|\boldsymbol{Y}} : \mathcal{B}^n \mapsto \{0, 1, 2, \cdots, M\}$ where 0 indicates that the decoder choose error. If messages are equiprobable, the average error probability is defined as $P_b = 1 - \frac{1}{M} P_{Z|\boldsymbol{X}}(m|\boldsymbol{X}_m)$. An codebook with M codewords and a decoder whose average probability of error is smaller than ϵ are called an (n, M, ϵ)-code.

An upper bound on the size of a code as a function of the average probability of symbol error follows the Corollary 1.

Theorem 7: Every (n, M, ϵ)-code for a random transformation $P_{Z|\boldsymbol{Y}}$ satisfies

$$\log M \leq \sup_{\boldsymbol{X}} I(\boldsymbol{X};\boldsymbol{Y}) - D(\boldsymbol{p}\|\boldsymbol{w}) + n\left(1 + P_s \log(q-1)\right) \tag{16}$$

where $\boldsymbol{p} = \{p_0, p_1, \cdots, p_n\}$ is the error distribution with $p_k = \Pr\left(H_d(\boldsymbol{X}, \boldsymbol{Y}) = k\right)$ for $k = 0, 1, \cdots, n$, $\boldsymbol{w} = \{w_0, w_1, \cdots, w_n\}$ is a probability distribution with $w_k = \frac{C_n^k}{2^n}$.

Proof:

Since the messages are equiprobable, we have $H(\boldsymbol{X}) = \log M$. According Corollary 1,

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) =& H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y}) \\ \geq& \log M - n + D(\boldsymbol{p}\|\boldsymbol{w}) - nP_s \log(q-1). \end{aligned} \tag{17}$$

Solving $\log M$ from (17), one can get (16), which completes the proof. ■

IV. Application to Channel Coding

Consider information transmission over a memoryless discrete q-ary symmetric channel with a channel code (n, M, ϵ) with crossover probability ε, as shown in Fig. 1. In this case, the probability of symbol error is $p_e = (q-1)\varepsilon$.

Then the error probabilities are

$$p_k = \Pr\{H_d(\boldsymbol{X}, \boldsymbol{Y}) = k\} = C_n^k p_e^k (1-p_e)^{n-k} \tag{18}$$

and the block error probability is

$$P_b = 1 - p_0 = 1 - (1-p_e)^n. \tag{19}$$

Using the extended Fano's inequality in Theorem 4, we have

$$H_e(\boldsymbol{X}|\boldsymbol{Y}) \leq \sum_{k=0}^{n} p_k \log\left(\frac{C_n^k (q-1)^k}{p_k}\right). \tag{20}$$

It is easy to see that the conditional entropy in theory is

$$\begin{aligned} H(\boldsymbol{X}|\boldsymbol{Y}) &= H\left(1-p_e, \varepsilon, \cdots, \varepsilon\right) \\ &= -\left(1-p_e\right)\log(1-p_e) - (q-1)\log\varepsilon. \end{aligned} \tag{21}$$

By Corollary 2, mutual information is lower bounded by

$$\begin{aligned} I_e(\boldsymbol{X};\boldsymbol{Y}) &\geq \sum_{k=0}^{n} p_k \log\left(\frac{C_n^k p_e^k (1-p_e)^{n-k}}{(C_n^k (q-1)^k)/q^n}\right) \\ &= n\log q + \sum_{k=0}^{n} p_k \left[k\log p_e + (n-k)\log\left(1-p_e\right)\right. \\ &\qquad \left. - k\log(q-1)\right]. \end{aligned} \tag{22}$$

while the capacity of the memoryless QSC is given by

$$\begin{aligned} I(X;Y) &= \log q - H\left(1-p_e, \varepsilon, \cdots, \varepsilon\right) \\ &= \log q + (1-p_e)\log(1-p_e) + (q-1)\varepsilon\log\varepsilon. \end{aligned} \tag{23}$$

And the relative entropy $D(\boldsymbol{p}\|\boldsymbol{w})$ can be derived as

$$\begin{aligned} D(\boldsymbol{p}\|\boldsymbol{w}) &= \sum_{k=0}^{n} p_k \log\left(\frac{C_n^k \varepsilon^k (1-(q-1)\varepsilon)^{n-k}}{C_n^k/2^n}\right) \\ &= n + \sum_{k=0}^{n} p_k \left[k\log p_e + (n-k)\log(1-p_e)\right] \end{aligned} \tag{24}$$

For a given average symbol error probability constraint $P_s = \epsilon$, the upper bound on the maximum codebook size given by Theorem 7 is

$$\begin{aligned} \log M_e \leq& I(\boldsymbol{X};\boldsymbol{Y}) - D(\boldsymbol{p}\|\boldsymbol{w}) + n\left(1 + P_s\log(q-1)\right) \\ =& n\log q - nH(1-p_e, \varepsilon, \cdots, \varepsilon) - \sum_{k=0}^{n} p_k \left[k\log p_e\right. \\ &\left. + (n-k)\log(1-p_e)\right] + n\epsilon\log(q-1). \end{aligned} \tag{25}$$

On the other hand, by Fano's inequality we have

$$H_f(\boldsymbol{X}|\boldsymbol{Y}) \leq H(P_b) + P_b\log(q^n - 1) \tag{26}$$

with P_b given by (19).

Then the lower bound of the mutual information is

$$I_f(\boldsymbol{X};\boldsymbol{Y}) \geq (1-P_b)\log q^n - H(P_b). \tag{27}$$

Finally, the upper bound on the codebook size is

$$\log M_f \leq \frac{1}{1-\epsilon}\left(nI(X;Y) + H(\epsilon)\right). \tag{28}$$

Suppose the QSC parameter are $\varepsilon = 0.001$ and $q = 7$, we calculated the bounds on conditional entropy, mutual information and codebook size by our proposed results and Fano's inequality.

Firstly, the upper bound on the conditional entropy is presented in Fig. 2. Specially, $H_e(\boldsymbol{X}|\boldsymbol{Y})$ is obtained by the extended Fano's inequality (20), $H_f(\boldsymbol{X}|\boldsymbol{Y})$ is calculated according to Fano's inequality (26) and $H(\boldsymbol{X}|\boldsymbol{Y})$ is the conditional entropy in theory (21). It is clear that Theorem 4 is tighter than Fano's inequality. Particularly, we have $H_e(\boldsymbol{X}|\boldsymbol{Y}) = H(\boldsymbol{X}|\boldsymbol{Y})$ for the QSC. This is because the error distributions are the same for any $\boldsymbol{Y} = \boldsymbol{y}$. So $H(E) = H(E|\boldsymbol{Y})$ holds. Besides, the error pattern is uniformly distributed for a given k, regardless of $\boldsymbol{y}$ and $H(E) = H(\boldsymbol{Y}|E, \boldsymbol{Y}) = H(\boldsymbol{Y}|E) = \log C_n^k (q-1)^k$ holds. Therefore, the upper bound is tight. However, for Fano's equality, there are relaxations in both $H(E) = H(E|\boldsymbol{Y})$ to $H(E)$ and $H(\boldsymbol{X}|E, \boldsymbol{Y})$ to $P_b\log(M-1)$.

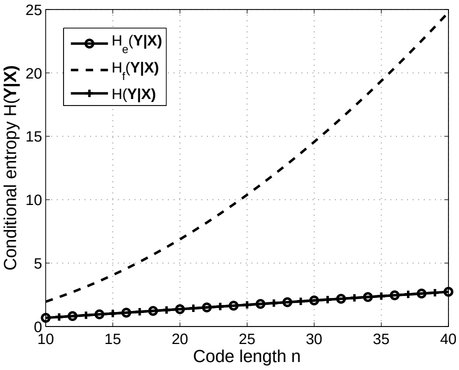

Fig. 2. The upper bound on conditional entropy, $q = 7, \varepsilon = 0.001$

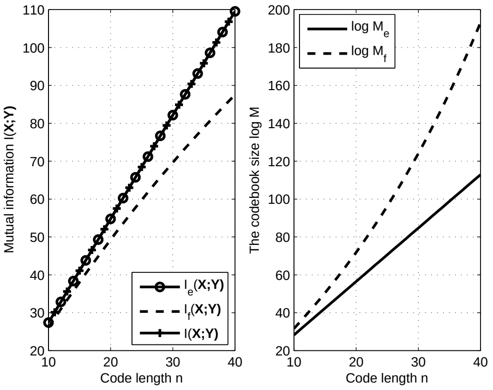

Fig. 3. Bounds on mutual information and codebook size v.s. blockleng, $q = 7, \varepsilon = 0.001$

Similarly, the lower bound on mutual information $I_e(\boldsymbol{X};\boldsymbol{Y})$ given by (22) coincides with $I(\boldsymbol{X};\boldsymbol{Y})$ in theory, given by (23) and is better than that given by Fano's inequality (27).

When we use the upper bound on the codebook size in Theorem 7, it should be noted that it is presented as a function of symbol error probability P_s. In fact, P_b is always larger than P_s. Therefore, we use the same fraction of them in the calculation of the bounds to make sense of the comparison, i.e., $\epsilon = \frac{P_s}{2}$ for (25) and $\epsilon = \frac{P_b}{2}$ for (28). It is also seen from Fig. 3 that our new developed result is tighter.

The performances of Theorem 6 and Theorem 7 versus the QSC parameter ε are shown in Fig. 4, where the block length is chosen as $n = 30$. As shown, the mutual information bound is tight and our results are much better than Fano's inequality. In the calculation of the upper bounds on codebook size, the selection of the error probability constraints are also chosen as $\epsilon_e = \frac{P_s(\varepsilon)}{2}$ and $\epsilon_f = \frac{P_b(\varepsilon)}{2}$ so that they are comparable.

V. Conclusion

In this paper, we revisited Fano's inequality and extended it to a general form. Particularly, the relation between the conditional entropy and error probability of two random vectors was considered, other than that between two random variables. This makes the developed results more suitable for source/channel codings in the finite blocklength regime. By investigating the block error pattern more detailedly, the conditional entropy of the original random vector given the received one is upper bounded more tightly by the extended Fano's inequality. Furthermore, the extended Fano's inequality is completely tight for some symmetric channels such the q-ary symmetric channels. Converse results are also presented in terms of lower bounds on the mutual information and a upper bound on the codebook size under the blocklength and symbol error constraints, which also have better performances.

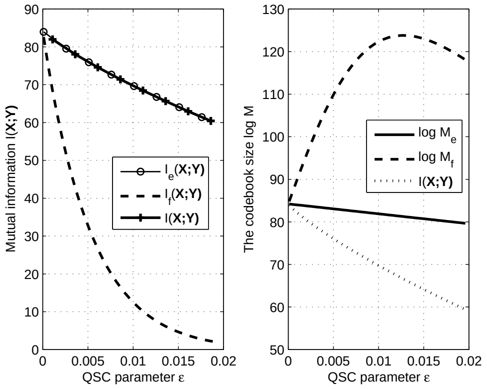

Fig. 4. Bounds on mutual information and codebook size v.s. QSC parameter ε, $q = 7, n = 30$

Acknowledgement

This work was partially supported by INC research grant of Chinese University of Hongkong and the China Major State Basic Research Development Program (973 Program) No. 2012CB316100(2).

References

[1] C. E. Shannon,"A mathematical theory of communication," *Bell Syst. Tech. J.* vol. pp.623-656, Oct. 1948.

[2] R. M. Fano, *Class Notes for Transmission of information*, Course6.574. Combridge, MA: MIT. 1952.

[3] R. W. Raymond, *Information Theory and Network Coding*, Berlin/Newyork: Springer. 2008.

[4] T.S. Han and S. Verdu, "Generalizing the fano inequality," *IEEE Trans. Inform. Theory*, vol. 40, no. 7, pp. 1247-1250, Jul. 1994.

[5] S. W. Ho and S. Verdu, "On the interplay between conditional entropy and error probability" *IEEE Trans. Inform. Theory*, vol.56, no.12, pp.5930-5942, Dec. 2010.

[6] D. L. Tebbe and S. J. Dwyer, III, "Uncertainty and probability of error," *IEEE Trans. Inform. Theory*, vol. 14, no. 3, pp. 516-518, May 1968.

[7] M. Feder and N. Merhav, "Relations between entropy and error probability," *IEEE Trans. Inform. Theory*, vol. 40, no. 1, pp. 259-266, Jan. 1994.

[8] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307-2359, May 2010.